Regression Analysis of Electrical Energy Consumption with Cross-Country data

Jun Xiang

Ziyu Liu

Youngjun Kim

ECON 3161

Georgia Institute of Technology

March 16 2016

Abstract:

This paper performs a cross-country analysis in order to reveal the relationship among energy consumption, country GDP, country energy net export and country population in developing and developed countries. After performing the linear regression model by using data from authoritative websites, a correlation between the independent variables and the dependent variable is confirmed.

Introduction:

Electrical energy is produced and consumed all the time at all countries. This paper performs a cross-country analysis on what affects annual energy consumption in 2012. Annual energy consumption is defined as the actual amount of energy generated by power plants. Since we want the analysis to be as comprehensive as possible, we want to make the sample size as large as possible, so the sample of the analysis is consisted by countries of different level: developing countries and developed countries, countries from different states, countries with different culture, etc. The goal of the paper is not only to determine the cause of energy consumption, but also to help understand the role of energy in economics. While country develops through time, power plant construction should follow the trend of development. The paper serves as a reference for policy makers to estimate extra power capacity needed according to GDP.

The hypothesis is that annual energy consumption has positive correlation with the country’s GDP, energy net export and population.

Literature Review:

Three journey papers are selected in the literature review. The paper, “*Relationship between Electricity Energy Consumption and GDP growth: Evidence from India*”, indicates the causality of energy consumption to GDP in India (Mohanty, 2015). The paper, "*Energy Consumption and GDP: Causality Relationship in G-7 Countries and Emerging Markets*", examines G-7 countries and concludes stationary linear relationship between energy consumption and GDP among all G-7 countries (Soytas, 2003). The report, “*Annual Energy Outlook 2015*” by U.S. Energy Administration, estimates energy consumption in a future period and implies a positive relationship between power capacity and energy consumption (Tubb 2015).

“*Relationship between Electricity Energy Consumption and GDP growth: Evidence from India*” takes the energy consumption and GDP data from 1950-1951 and 1996-1997 in India to examine the causality and direction of the two variables. Its examination indicates causality of energy consumption to GDP growth and emphasizes the importance of the policy of “energy must lead economic growth”.

"*Energy Consumption and GDP: Causality Relationship in G-7 Countries and Emerging Markets*" performs less analysis than the first paper, but extracts and compresses information from other studies. It examines the same causality and direction as in the first paper, but in G-7 countries. Stationary linear relationship as well as causality is observed for all these countries, but the direction of causality is different for different countries.

The first two papers lie the fundamental confidence of finding positive correlation between energy consumption and GDP. For power generation capacity, most journal article focuses on doing forecasting following the underlying logic that the power plant construction should follow growth of energy demand. The paper takes such logic and assumes such forecasting is achievable. Thus the quantitative relationship of power capacity and energy consumption can be determined. Under credibility concern, this paper refers to the report, “*Annual Energy Outlook 2015*” by U.S. Energy Administration. The report estimates energy consumption in a future period and implies a positive relationship between power capacity and energy consumption.

In our paper, we took the inspiration from the above literature, and we decided to take a step further. We not only put GDP into the consideration, but also added more variables to see if there exist some correlation between energy consumption and the variables.

Data:

The dependent variable is obviously the energy consumption of the country. What factors might have a relationship with the energy consumption? As stated in the literature review, it is not hard to think that one factor is the “strength” of the country: namely, the country’s GDP. The most developed countries have busiest transportation and heaviest industry and agriculture, and it would make sense if higher GDP implies higher energy consumption. By similar argument, it would also make sense if lower GDP implies lower energy consumption. Another factor is the net export of the country, which is similar to the GDP factor, since high net export means that the economy of the country is in a good position, and thus a higher energy consumption. The third factor is population. Intuitively, this factor should also have some correlation with energy consumption, because higher population means more transportation, more use of electricity, etc. Thus, we have three independent variables: GDP of the country, net export of the country and population of the country.

The data of each country’s energy consumption and net export (which includes the data of import and export) are collected from US Energy Information Administration. The data of each country’s GDP and population are collected from the most authoritative website: The World Bank. Both websites provide open access to data about development in countries around the globe, and they are updated regularly, so the data is absolutely reliable.

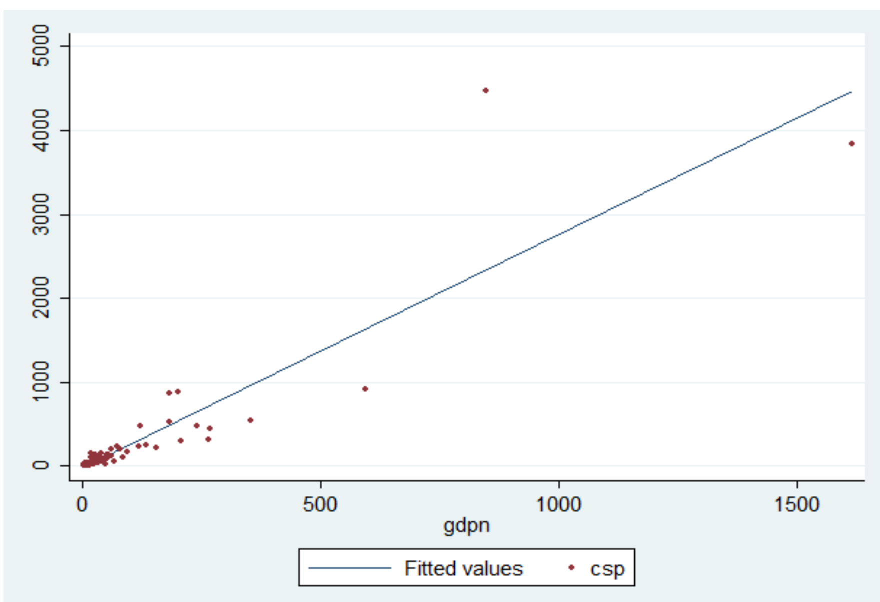

Figure 1. Relationship between csp and gdpn

Figure 1 is a scatter plot that shows the relationship between consumption (in billion kilowatthours) and GDP (in 10^10 USD). The x-axis is the independent variable, gdpn, and the y-axis is the dependent variable, csp. It contains all 140 samples, and it is not hard to see that the graph contains a cluster at the bottom left corner, and a few outliers that are very far away from the cluster. To have a clearer view of the majority of the data, we take out 14 samples with highest GDP and draw "zoomed-in" version of Figure 1.

‘

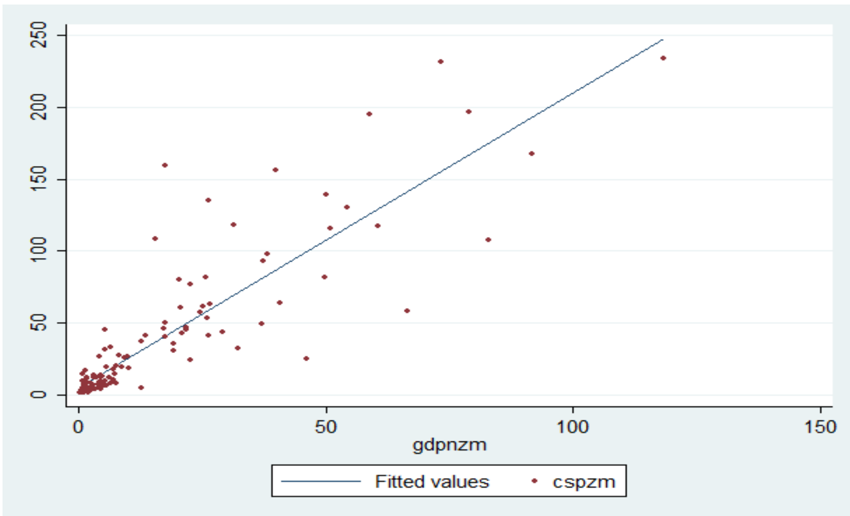

Figure 2. Relationship between cspzm and gdpnzm

Similar as Figure 1, Figure 2 is another scatter plot that shows the relationship between consumption (in billion kilowatthours) and GDP (in 10^10 USD), but gives a better view of the data with 126 samples. The slope of the line from both Figure 1 and Figure 2 look very positive, which indicates that GDP and energy consumption are highly related. More specific data and analyzation of the graph will be presented in the Result section.

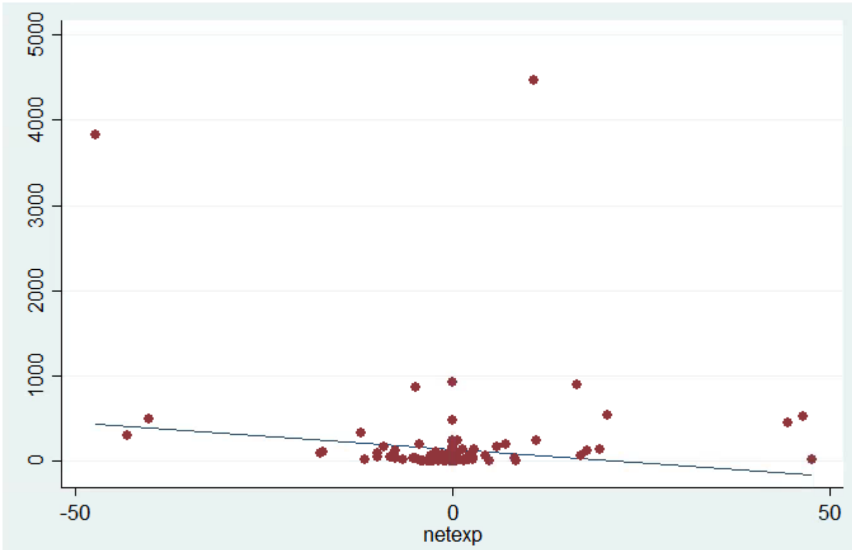

Figure 3. Relationship between csp and netexp

Figure 3 is a scatter plot that shows the relationship between consumption (in billion kilowatthours) and net export (in billion kilowatthours). The x-axis is the independent variable, netexp, and the y-axis is the dependent variable, csp. It contains all 140 samples, and surprisingly, there seem not to be a strong relationship between energy consumption and energy net export. More details will be presented in the Result section.

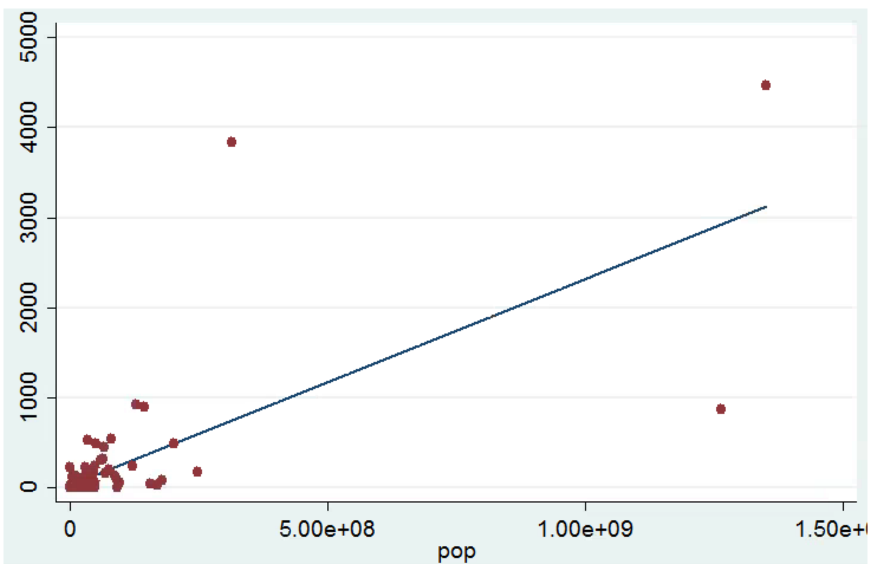

Figure 4. Relationship between csp and pop

Figure 4 is a scatter plot that shows the relationship between consumption (in billion kilowatthours) and population. The x-axis is the independent variable, population, and the y-axis is the dependent variable, csp. It contains all 140 samples, and unsurprisingly, there seem to be a positive relationship between consumption and population. Figure 5 below is similar as Figure 2, a zoom-in version of Figure 4 which gives a better view of data.

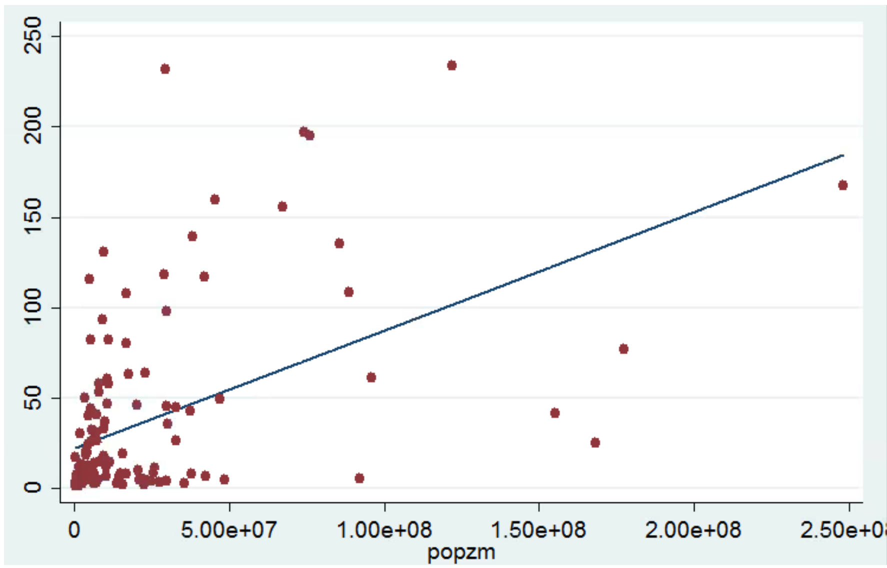

Figure 5. Relationship between cspzm and popzm

All of the Gauss Markov Assumptions are met in our model. $y = \beta 0 + \beta 1x1 + \beta 2x3 + \beta 3\ x3 + u$ satisfies linear parameter. The samples are randomly picked, so it satisfies random sampling. Since none of the independent variable is constant and there's no linear relationship among independent variables, there's no perfect collinearity. The error u has an expected value of 0 given any values of independent variables, so it satisfies zero condition means. Finally, the error *u* has the same variance given any values of the explanatory variables, so it satisfies homoscedasticity.

Results:

Dependent Variable: Electrical Energy Consumption				
Independent Variables	Number of samples	Beta	Intercept	R^2
GDP	140	2.77e-10	-7.10	0.824
Net Export	140	-6.18	136.68	0.017
Population	140	2.30e-6	23.84	0.514

Table 1. Single variable regression

All of the data that is in the following three tables is from the STATA regression, as shown in the table 4-8 in the Appendix. Table 1 shows the single variable regression between energy consumption and each independent variable. GDP has the highest R^2 value, which is 0.824. This means that GDP is highly correlated with energy consumption, and the positive beta means that they are positively correlated. For every increase of 10^10 dollars, an increase of 2.77 billion kilowatthours is expected. Population has an R^2 value of 0.514, which means that it is fairly correlated with energy consumption. Similar as GDP, population has a positive beta value and thus is positively correlated with energy consumption. For every increase of 10^6 population, an increase of 2.30 billion kilowatthours is expected. Both GDP and population's R^2 values are within expectation. However, R^2 for net export turns out to be a surprise. It is only 0.017, which is extremely low. This means that there's almost no correlation between the independent variable, net export, and the dependent variable, energy consumption. This is totally different from what we were expecting at the very beginning.

Dependent Variable: Electrical Energy Consumption			
Independent Variables	Model 1	Model 2	Model 3
GDP	T-stat=25.37	T-stat=25.26	T-stat=25.37
Net Export	---	---	T-stat=2.22
Population	---	T-stat=12.01	T-stat=11.92
Intercept	T-stat=-0.37	T-stat=-2.39	T-stat=-2.53
R^2	0.8235	0.9100	0.9170
Number of Observations	140	140	140

Table 2. Multivariable Regression with different models

Table 2 presents three models, where model 1 is a single variable regression of GDP, model 2 is a multivariable regression of GDP and population, model 3 is a multivariable regression of all three independent variables. The T-stat is provided for each model, and we will focus on model 2 and model 3. We now conduct the hypothesis test. Hypothesis 1: Country electrical consumption is correlated with country GDP, energy net export and population. Hypothesis 2: Country electrical consumption is correlated with country GDP and population but not with net export.

Dependent Variable: Electrical Energy Consumption		
Independent Variables	Model 2	Model 3
GDP	P(X>|T|)=0	P(X>|T|)=0
Net Export	---	P(X>|T|)=0.028
Population	P(X>|T|)=0	P(X>|T|)=0
Intercept	P(X>|T|)=0.018	P(X>|T|)=0.013

Table 3. Hypothesis test table

As seen above from Table 3 in Model 2 and 3, the regression results for GDP population has zero probability to be false. The result for export, however, has a probability of 0.028 to be false. This is too small to reject the first hypothesis. Therefore, we reject hypothesis 2 but failed to reject hypothesis 1. As a result,

Conclusion:

The initial hypothesis is correct: a country's annual energy consumption has positive correlation with the country's GDP, energy net export and population. The final equation we get is $y = -33.81 + 2.28e\text{-}10\ gdp + 1.10e\text{-}6\ pop + 2.69\ netexp$. From this equation, we can roughly predict energy consumption given a country's GDP, population and net export, and we can also forecast power plants construction.

Now the question is, can we improve the model to make it more accurate? Like most models, ours is not a perfect one. There are other potential variables that affect energy consumption. Actually, we are still trying to add some other factor that might be relevant, such as the culture background of the country: social system, religion, etc. However, the difficulty is to figure out how to quantify them. In addition, many countries practice more than one religions, which makes it even more difficult to analyze. Hopefully in the future, we can find a way to add these factors to the analyzation, and make our model one step closer to the perfect one.

References

1.Mohanty, Asit, and Devtosh Chaturvedi. "Relationship between Electricity Energy Consumption and GDP: Evidence from India." International Journal of Economics and Finance IJEF 7.2 (2015). Print.

2.Soytas, Ugur, and Ramazan Sari. "Energy Consumption and GDP: Causality Relationship in G-7 Countries and Emerging Markets." Energy Economics 25.1 (2003): 33-37. Print.

3.Tubb, Rita. "EIA: Annual Energy Outlook through 2040." Pipeline & Gas Journal 1 July 2015. Print.

4."International Energy Statistics - EIA." International Energy Statistics - EIA. Web. Feb 25th. 2016.**http://www.eia.gov/cfapps/ipdbproject/iedindex3.cfm?tid=2&pid=2&aid=2&cid=regions&syid=2012&eyid=2012&unit=BKWH**

5. “The World Bank”. Web. Feb 25th. 2016. http://data.worldbank.org/indicator/SP.POP.TOTL

Appendix

```
. regress csp gdpn
```

Source	SS	df	MS
Model	30004135.1	1	30004135.1
Residual	6431087.13	138	46602.0807
Total	36435222.2	139	262123.901

Number of obs	=	140
F(1, 138)	=	643.84
Prob > F	=	0.0000
R-squared	=	0.8235
Adj R-squared	=	0.8222
Root MSE	=	215.88

csp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdpn	2.765883	.1090049	25.37	0.000	2.550347	2.981418
_cons	-7.096073	19.10896	-0.37	0.711	-44.88029	30.68814

Table 4. Single variable: regress with GDP

```
. regress csp netexp
```

Source	SS	df	MS
Model	616084.347	1	616084.347
Residual	35819137.9	138	259558.97
Total	36435222.2	139	262123.901

Number of obs	=	140
F(1, 138)	=	2.37
Prob > F	=	0.1257
R-squared	=	0.0169
Adj R-squared	=	0.0098
Root MSE	=	509.47

csp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
netexp	-6.189444	4.017442	-1.54	0.126	-14.13315	1.754258
_cons	136.6818	43.05874	3.17	0.002	51.54155	221.822

Table 5. Single variable: regress with net export

```
. regress csp pop

      Source |       SS           df       MS      Number of obs   =       140
-------------+----------------------------------   F(1, 138)       =    145.71
       Model |  18712415.2         1  18712415.2   Prob > F        =    0.0000
    Residual |    17722807       138  128426.138   R-squared       =    0.5136
-------------+----------------------------------   Adj R-squared   =    0.5101
       Total |  36435222.2       139  262123.901   Root MSE        =    358.37

------------------------------------------------------------------------------
         csp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         pop |   2.30e-06   1.90e-07    12.07   0.000     1.92e-06    2.67e-06
       _cons |   27.83612   31.61041     0.88   0.380    -34.66725    90.33949
------------------------------------------------------------------------------
```

Table 6. Single variable: regress with population

```
. regress csp gdp pop

      Source |       SS           df       MS      Number of obs   =       140
-------------+----------------------------------   F(2, 137)       =    727.97
       Model |  33301648.7         2  16650824.4   Prob > F        =    0.0000
    Residual |  3133573.47       137  22872.7991   R-squared       =    0.9140
-------------+----------------------------------   Adj R-squared   =    0.9127
       Total |  36435222.2       139  262123.901   Root MSE        =    151.24

------------------------------------------------------------------------------
         csp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |   2.23e-10   8.84e-12    25.26   0.000     2.06e-10    2.41e-10
         pop |   1.12e-06   9.29e-08    12.01   0.000     9.32e-07    1.30e-06
       _cons |  -32.32643   13.55124    -2.39   0.018    -59.12308   -5.529786
------------------------------------------------------------------------------
```

Table 7. Multivariable – model 2: regress with GDP and population

```
. regress csp gdp pop netexp
```

Source	SS	df	MS			
				Number of obs	=	140
				F(3, 136)	=	500.81
Model	33410859.1	3	11136953	Prob > F	=	0.0000
Residual	3024363.09	136	22237.9639	R-squared	=	0.9170
				Adj R-squared	=	0.9152
Total	36435222.2	139	262123.901	Root MSE	=	149.12

csp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdp	2.28e-10	8.99e-12	25.37	0.000	2.10e-10	2.46e-10
pop	1.10e-06	9.20e-08	11.92	0.000	9.15e-07	1.28e-06
netexp	2.690546	1.214104	2.22	0.028	.2895812	5.091511
_cons	-33.81039	13.37863	-2.53	0.013	-60.26744	-7.353333

Table 8. Multivariable – model 3: regress with all three independent variables